The first multi-tissue diel genome-scale metabolic model of a woody plant highlights suberin biosynthesis pathways in *Quercus suber*

Emanuel Cunha[1,2], Miguel Silva[1,2], Ines Chaves[3;4], Huseyin Demirci[1,2,5], Davide Rafael Lagoa[1,2], Diogo Lima[1,2], Miguel Rocha[1,2,a], Isabel Rocha[3,a,b], and Oscar Dias[1,2,a,b]

[1] Centre of Biological Engineering, Universidade do Minho, 4710-057 Braga, Portugal

[2] LABBELS – Associate Laboratory, Braga, Guimarães, Portugal

[3] Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Avenida da República, Quinta do Marquês, 2780-157 Oeiras, Portugal

[4] iBET, Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras, Portugal

[5] SnT/University of Luxembourg, Luxembourg

[a] Senior authors

[b] Corresponding authors: irocha@itqb.unl.pt; odias@ceb.uminho.pt

Abstract

In the last decade, genome-scale metabolic models have been increasingly used to study plant metabolic behavior at the tissue and multi-tissue level under different environmental conditions. *Quercus suber*, also known as the cork oak tree, is one of the most important forest communities of the Mediterranean/Iberian region. In this work, we present the genome-scale metabolic model of the *Q. suber* (iEC7871), the first of a woody plant. The metabolic model comprises 7871 genes, 6231 reactions, and 6481 metabolites across eight compartments. Transcriptomics data was integrated into the model to obtain tissue-specific models for the leaf, inner bark, and phellogen, with specific biomass compositions. The tissue-specific models were merged into a diel multi-tissue metabolic model to predict interactions among the three tissues at the light and dark phases. The metabolic models were also used to analyze the pathways associated with the synthesis of suberin monomers. Nevertheless, the models developed in this work can provide insights into other aspects of the metabolism of *Q. suber*, such as its secondary metabolism and cork formation.

Keywords: Cork Biosynthesis, *Quercus suber* (Cork Oak), Genome-scale Metabolic Model, Multi-tissue diel cycle model, Secondary Metabolism, Suberin

1. Introduction

The cork oak, *Quercus suber* L., is a characteristic tree of the Mediterranean/Iberian landscape ecosystem. The tree forms a thick bark of cork (phellem) containing high levels of aliphatic suberin,

aromatic suberin, extractives (waxes and tannins), and polysaccharides (Graça, 2015). Cork properties are so unique that this material is used in diverse applications from the common bottle-stopers to spacecrafts built by NASA and European Space Agency (ESA), providing insulation solutions to deal with extreme conditions. The aliphatic suberin component has particular interest since it provides waterproof, light, elastic, and fire-retardant properties to cork (Vaz *et al.*, 2011; Pereira, 2015). Only the outer bark is separated from the trunk during the harvest, which enables the regeneration and allows using the tree as a renewable biological resource. Cork extraction can continue in the same tree for more than a century every 9 years. A cork oak tree with 234 years old, debarked 20 times, is now a living icon being elected European tree of the year 2018 (https://www.treeoftheyear.org/Previous-Years/2018).

Phellem is produced by the phellogen (cork cambium), involving the proliferation of phellogen derivates cells which undergo to differentiate into cork cells through cell expansion, deposition of suberin and waxes, and an irreversible senescence program ending with cell death. The first debarking occurs when the tree has 18-25 years old. This first harvested cork, known as “virgin” cork, produced from the original phellogen, yields poor-quality cork. The original phellogen is then replaced by a traumatic phellogen that proliferates to originate a new cork layer. From the third debarking onwards, a higher quality cork with high economic value is obtained - reproduction cork (or “amadia”) (Graça & Pereira, 2004). Cork growth and quality are dependent on the genome of each tree but it is also strongly dependent on the environment, such as water availability, temperature, pests, and diseases.

The cork oak tree represents one of the most relevant broadleaved forest resources in the Mediterranean basin (Portugal, Spain, Algeria, Morocco, France, Italy, and Tunisia), having a substantial socio-economic and ecological impact on these countries. This tree species can generally live over 200 years, which implies a capacity to bear with biotic stresses like bacteria, fungi and insects, as well as abiotic stresses like drought, floods and fires (Acácio *et al.*, 2007; Kim *et al.*, 2017).

Due to its economic interest in Portugal, a national consortium was created to first sequence the transcriptome (Pereira-Leal *et al.*, 2014) and recently to full sequence the genome of *Quercus suber* (Ramos *et al.*, 2018). The genome sequence's availability allows the application of systems biology tools to study this species' metabolic behavior.

Genome-scale metabolic (GSM) models aim at depicting the whole metabolic network of an organism. Such models have been widely used for metabolic engineering purposes, mainly with prokaryotes and yeasts. Among other applications, GSM models can be used to analyze an organism's metabolic behavior in different environmental and genetic conditions, including the effect of gene knock-outs and over/under-expressions (Dias & Rocha, 2015). Curated models have proven to accurately predict

complex organisms' metabolic behavior in diverse areas of knowledge, from biotechnological or environmental to medical applications (Fong *et al.*, 2005; Bordbar *et al.*, 2011; Agren *et al.*, 2014; Jerby-Arnon *et al.*, 2014; Zhang & Hua, 2016). Genome-scale modelling in plants is much more challenging than in prokaryotes. The struggle starts in the annotation of the complex genomic content, in which the function of a significant number of genes is still unknown. Also, the subcellular compartmentalization, tissue differentiation, and interactions between tissues are complex in plants. Nevertheless, in the last decade, the number of published GSM models of plants has increased considerably. Organisms, like *Arabidopsis thaliana* (Dal'Molin *et al.*, 2010; Cheung *et al.*, 2013)*, Zea mays* (maize) (Saha *et al.*, 2011; Seaver *et al.*, 2015), *Oryza sativa* (rice) (Poolman *et al.*, 2013; Lakshmanan *et al.*, 2013) have more than one model available. Considering the importance of plants in terms of nutrition, biofuels, and their capability to produce a variety of secondary metabolites, it is not surprising that studies for plant genome-scale model reconstruction will increase, parallel to the growth in the number of sequenced species. The utilization of user-friendly tools designed for this purpose, like *metabolic models reconstruction using genome-scale information* (*merlin)* (Dias *et al.*, 2015), accelerates the reconstruction process.

This paper describes the first reconstruction of iEC7871, a Genome-scale metabolic model for the cork oak tree, and as far as we know, the first published GSM model of a tree. Besides the generic GSM, we present tissue-specific models and a multi-tissue metabolic model that can be used to study the metabolic behavior of *Q. suber*.

2. Results

2.1. Genome Annotation

A GSM model for *Q. suber* based on an up-to-date genome annotation was reconstructed in this work. A Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) search against the Swiss-ProtKB (Apweiler *et al.*, 2011) database allowed to identify similarity results for 47,199 out of the 59,614 genes available in the genome (Ramos *et al.*, 2018). A second BLAST search against UniProtKB/TrEMBL allowed obtaining hits for 12,415 genes, while 590 had no results.

As detailed in the Materials and Methods section, *merlin*'s "automatic workflow" tool annotates genes based on the homologous gene records taxonomy. Most enzyme-encoding genes were annotated based on *A. thaliana* gene records (82%). Organisms not included in the annotation workflow represent 16% of gene annotations, and only 1% were annotated using *Quercus* genus' gene records. A more detailed analysis of the genome annotation is available in Supplementary File 1.

2.2. Model Properties

The metabolic model is mass balanced and can predict growth in phototrophic and heterotrophic conditions. These conditions were defined by setting the photon and CO_2 (phototrophic) or sucrose (heterotrophic) as the sole energy and carbon sources, respectively. Additionally, the model requires NH_3 or HNO_3, H_3PO_4, H_2SO_4, Fe^{2+}, and Mg^{2+} to produce biomass.

The general properties of the cork oak model and other five published plant models – *A. thaliana* (Chung *et al.*, 2013; Seaver *et al.*, 2015), *Z. mays* (Saha *et al.*, 2011; Seaver *et al.*, 2015), and *Glycine max* (soybean) (Moreira *et al.*, 2019) - are presented in table 1.

Table 1. General properties (genes, reactions, metabolites, and compartments) of the Q. suber model, and other five published plant GSM models. Only generic models were considered here.

	***Quercus suber* (this work)**	***Arabidopsis thaliana*** (Chung *et al.*, 2013)	***Arabidopsis thaliana*** (Seaver *et al.*, 2015)	***Zea mays*** (Saha *et al.*, 2011)	***Zea mays*** (Seaver *et al.*, 2015)	***Glycine max*** (Moreira *et al.*, 2019)
Genes	7871	1475	5184	1563	13,279	6127
Reactions	6231	1895	6399	1970	6458	2984
Metabolites	6481	1761	6236	2129	6250	2814
Compartments	8	6	9	6	9	5

The Cork Oak model comprises 7871 genes, 6231 reactions (enzymatic, spontaneous, and transport), 708 exchange reactions, and 6476 metabolites distributed across eight subcellular compartments – the cytoplasm, mitochondria, plastid, endoplasmic reticulum, Golgi apparatus, vacuole, peroxisome, and the extracellular space. The number of genes, reactions, and metabolites of the Cork Oak model are only comparable with the *A. thaliana* and *Z. mays* models developed by Seaver (Seaver *et al.*, 2015).

The reactions and metabolites present in the cork oak model were compared with the ones present in the models developed by Chung for *A. thaliana* (Chung et al., 2013), and Saha for *Z. mays* (Saha *et al.*, 2011) (Supplementary File 2), as the reactions and metabolites available in these models have the same identifiers (KEGG (Kanehisa & Goto, 2000)) as our model.

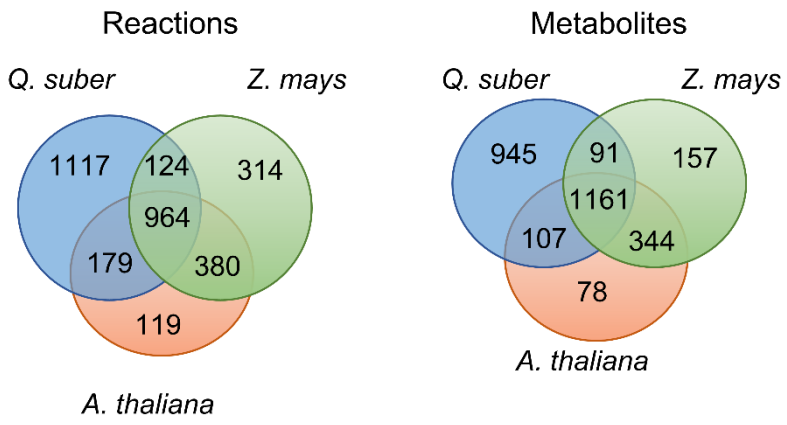


Figure 1. Venn diagram of reactions (left) and metabolites (right) included in the *Q. suber*, *A. thaliana*, and *Z. mays* models.

A total of 3196 unique reactions and 2883 unique metabolites were identified across the three models. The cork oak model includes 1117 reactions that are not present in the other two - Figure 1. It has 124 reactions in common only with the Maize model and 179 only with the *A. thaliana* model. The maize and *A. thaliana* models share 380 reactions, and 964 were found in the three models. The metabolites analysis shows similar behavior. This analysis is detailed in Supplementary File 2 Tables S5-S6.

The reactions only available in the Cork Oak model (called unique reactions in the following) were further analyzed to assess the annotation of the genes and pathways associated with them. This analysis is detailed in Supplementary File 2 Table S7. In agreement with the remaining genome annotation, 82% of the genes encoding enzymes that catalyze these reactions were annotated based on *A. thaliana* genes' annotations. The remaining 18% were annotated by gene records of organisms not accounted for in the annotation workflow. KEGG pathways with more unique reactions are presented in Table 2. The 34 pathways with at least 10 unique reactions can be divided into 7 main areas of the KEGG metabolism: 8 pathways representing 'Lipid metabolism', 7 the 'aminoacid metabolism' and 'carbohydrate metabolism', 6 the 'Xenobiotics biodegradation metabolism', 3 the 'Metabolism of cofactors and vitamins' and 'Biosynthesis of other secondary metabolites' and 1 the 'Metabolism of terpenoids and polyketides'. Spontaneous reactions and reactions not associated with any pathway were not considered.

139 Table 2. Number of reactions included in the cork oak model that were not identified in the *A. thaliana* (Cheung *et al.*, 2013)
140 and *Z. mays* (Saha *et al.*, 2011) models for each pathway (so-called unique reactions). Only pathways with more than 10
141 unique reactions were included in this table. The complete table is available in Supplementary File 2 S7. The KEGG pathways
142 were organized according to the 7 main areas of KEGG metabolism: 1- Xenobiotics biodegradation and metabolism; 2- Lipid
143 metabolism; 3- Metabolism of cofactors and vitamins; 4- Amino acid metabolism and Metabolism of other amino acids; 5-
144 Metabolism of terpenoids and polyketides; 6- carbohydrate metabolism; 7- Biosynthesis of other secondary metabolites

	KEGG Pathway	Number of unique reactions
1	Metabolism of xenobiotics by cytochrome P450	38
2	Steroid hormone biosynthesis	38
3	Ubiquinone and other terpenoid-quinone biosynthesis	34
4	Tryptophan metabolism	32
1	Steroid biosynthesis	24
4	Tyrosine metabolism	23
4	Cysteine and methionine metabolism	22
1	Drug metabolism - cytochrome P450	21
2	Fatty acid metabolism	21
2	Glycerophospholipid metabolism	20
2	Inositol phosphate metabolism	19
5	Brassinosteroid biosynthesis	17
2	Cutin, suberine and wax biosynthesis	17
1	Drug metabolism - other enzymes	17
4,6	Amino sugar and nucleotide sugar metabolism	16
3	Porphyrin and chlorophyll metabolism	16
2	Arachidonic acid metabolism	16
4	Arginine and proline metabolism	16
3	Nicotinate and nicotinamide metabolism	15
4	Cyanoamino acid metabolism	14
6	Butanoate metabolism	14
7	Betalain biosynthesis	14
1	Benzoate degradation	13
6	Fructose and mannose metabolism	11
6	Glyoxylate and dicarboxylate metabolism	11
6	Ascorbate and aldarate metabolism	11
2	Sphingolipid metabolism	11
7	Phenylpropanoid biosynthesis	10
7	Flavonoid biosynthesis	10
6	Propanoate metabolism	10
2	Fatty acid biosynthesis	10
6	Galactose metabolism	10
1	Chlorocyclohexane and chlorobenzene degradation	10
4	Lysine biosynthesis	10

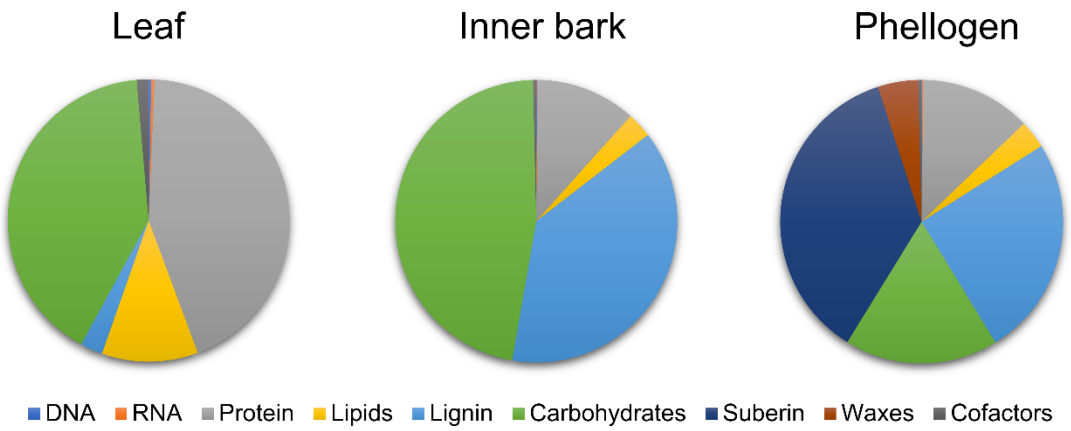
145 The unique reactions identified are associated with 122 different KEGG pathways, from which 54 had
146 five or fewer unique reactions. “Steroid hormone biosynthesis”, “Metabolism of xenobiotics by
147 cytochrome P450”, “Ubiquinone and other terpenoid-quinone biosynthesis”, and “Tryptophan
148 metabolism” are the pathways with more unique reactions, having 32 to 38 reactions unavailable in
149 the maize and *A. thaliana* models.

150 Several of these reactions can be related to species-specific reactions reflecting the cork oak
151 uniqueness and differences from non-woody plants. Although the reactions are cork oak specific, they
152 belong to pathways already present in other species. Some unique reactions are associated with
153 pathways essential for trees and belong to the secondary metabolism responsible for wood and cork
154 production, such as “Phenylpropanoid biosynthesis”, “Flavonoid biosynthesis”, “Fatty acid
155 biosynthesis”, “Metabolism of xenobiotics by cytochrome P450”, and “Cutin, suberine and wax
156 biosynthesis”.

157 ## 2.3. Tissue-specific models

158 Transcriptomics data were integrated into the generic model using troppo (Ferreira *et al.*, 2020) to
159 obtain tissue-specific models. Leaf, Inner bark, Reproduction or traumatic Phellogen, and Virgin
160 Phellogen were selected because of their influence on tree growth and cork production.

161 Different biomass formulations for each tissue (Figure 2) were determined, according to experimental
162 data available in the literature and other plant GSM models. The detailed biomass composition for
163 each tissue is available in Supplementary File 3.



164

165 Figure 2. Biomass composition of the leaf, inner bark, and phellogen determined using data retrieved from published plant
166 GSM models and available experimental data. The biomass composition and the respective data sources are detailed in
167 Supplementary File 3.

168 As explained in the Materials and Methods section, the biomass was formulated by creating “e-
169 Metabolites”, representing the macromolecular composition of each tissue. Each macromolecule (e.g.

170 e-RNA) is associated with a reaction responsible for producing it from its precursors (e.g. ATP, GTP,
171 CTP, UTP).

172 The leaf macromolecular contents were determined using *A. thaliana* models (Dal'Molin *et al.*, 2010;
173 Arnold & Nikoloski, 2014). The cell wall sugar content was included in the e-Carbohydrate
174 composition, while lignin was included in the e-Lignin composition. The monomer contents of DNA,
175 RNA, and protein were determined using the biomass tool available in *merlin*. The fatty acid, lipid, and
176 carbohydrate compositions were determined using experimental data for *Q. suber* or closely related
177 organisms when species-specific data was not available (Koiwai *et al.*, 1983; Nouairi *et al.*, 2006;
178 Passarinho *et al.*, 2006). The lignin, carbohydrate, suberin, and wax contents and composition in the
179 inner bark and phellogen were determined using available experimental data (Pereira, 1988; Lourenço
180 *et al.*, 2016).

181 The Cofactors component includes a set of universal cofactors and vitamins (Xavier *et al.*, 2017). These
182 compounds were included in the biomass of each tissue. Nevertheless, the leaf "e-Cofactor" reaction
183 also comprises a set of pigments, such as chlorophylls and carotenoids, determined according to
184 experimental data (Garcia-Plazaola, 1997).

185 The formulation of tissue-specific biomass composition is critical to obtain tissue-specific models and
186 predict each tissue's metabolic behavior and the interactions among them. The leaf is mainly
187 composed of carbohydrates and protein. The inner bark also presents high amounts of carbohydrates
188 and a significant content of lignin. The phellogen also exhibits considerable amounts of these
189 macromolecules, but suberin is the one with the highest representation.

190 For tissue-specific model construction, reactions encoded by genes identified as "not expressed" were
191 removed from each tissue-specific model. The number of genes and reactions decreased in rather
192 different proportions, as *troppo* operates at the reaction level (See Material and Methods section 4.3).
193 Table 3 shows the number of reactions and metabolites in the generic model and each tissue-specific
194 model.

195

196

197

198

199

200 Table 3. General properties of the generic and tissue-specific GSM models. Genes were predicted based on the cork oak
201 genome and then used to develop the GSM model. The reactions were divided according with the respective metabolic role:
202 metabolic (enzymatic and spontaneous) and transport, and the respective compartment. The number of genes, metabolites,
203 and reactions were determined in the generic model and in each tissue-specific model, generated with *troppo*.

Component	Generic Model	Leaf	Inner Bark	Phellogen (reproduction)	Phellogen (virgin)
Genes	7871	7125	7004	7069	7044
Metabolites	6481	5222	5056	5239	5226
Exchange reactions	708	574	545	590	577
Reactions (metabolic)	6231	4514	4533	4589	4516
• **Enzymatic/ Spontaneous**	4961	3528	3589	3572	3554
• **Transport**	1270	986	944	1017	962
• **Extracellular**	176	119	120	142	140
• **Plastid**	1355	984	939	947	934
• **Cytoplasmic**	1644	1151	1177	1171	1172
• **Mitochondria**	722	532	568	517	518
• **Peroxisome**	296	219	220	206	205
• **Vacuole**	130	87	76	93	90
• **Endoplasmic Reticulum**	583	391	443	453	451
• **Golgi Apparatus**	55	45	45	43	44

204

205 The number of metabolic reactions is quite similar among the four tissue-specific models.
206 Nevertheless, the number of reactions in the leaf model's plastid is slightly higher than in the other
207 models, while in the endoplasmic reticulum, the opposite behavior can be observed. The reproduction
208 phellogen model presents a higher number of transport reactions than the remaining ones. To assess
209 potential differences in the metabolism of the reproduction and virgin phellogen, *in silico* simulations
210 with parsimonious flux balance analysis (pFBA) (Lewis *et al.*, 2010) and flux variability analysis (FVA)
211 were performed (Supplementary File 4 – Table S1).

212 The number of metabolic reactions associated with each pathway is available in Supplementary File 2
213 Tables S8-S9. The number of reactions associated with the "Arachidonic acid metabolism",
214 "Brassinosteroid biosynthesis", and "Drug metabolism – cytochrome P450" pathways are higher in the
215 inner bark and phellogen than in the leaf. The "Cutin, suberin and wax biosynthesis" pathway is mostly
216 represented in the phellogen. As expected, the "Carotenoid biosynthesis" pathway is present in the
217 leaf, while in the remaining tissues, the number of reactions is more limited. No significant differences
218 were identified between the number of reactions in each pathway between the reproduction and
219 virgin phellogen models.

220 2.4. Model Validation

221 The models developed in this work were tested to guarantee that no biomass or energy is produced
222 without energy input in each condition. The general and tissue-specific metabolic models were
223 validated by analyzing the fluxes of *in silico* simulations in photoautotrophic, heterotrophic, and
224 photorespiratory conditions. The results for model's validation are available in detail in Supplementary
225 File 5 Tables S1-S6 and Fig. S1-S2. All models are able to grow in heterotrophic conditions, while in
226 photoautotrophic and photorespiratory conditions only the general and leaf models produce biomass.

227 As expected, the photoautotrophic growth is supported by the assimilation of CO_2 by RubisCO using
228 light as energy source. The predicted photon uptake flux was 53.68 $mmol \cdot gDW^{-1} \cdot h^{-1}$ for a
229 biomass production fixed at 0.11 h^{-1} (see Materials and Methods). In heterotrophic conditions using
230 sucrose as carbon source, the tricarboxylic cycle (TCA) and the oxidative phosphorylation become
231 more active providing energy to sustain growth. The metabolic response to photorespiration was also
232 assessed in the leaf model. In these conditions, the 'Glyoxylate and dicarboxylate metabolism'
233 pathway, which includes the reactions associated with photorespiration, becomes more active to
234 recycle carbon skeletons (Supplementary File 5 Fig. S1).

235 Quantum yield (amount of CO_2 fixed per mol of photons) and assimilation quotient (CO_2 fixed per O_2
236 evolved) are common measures of photosynthetic efficiency. Thus, these parameters were
237 determined for the leaf model. In photoautotrophic conditions, the photon yield was calculated as
238 0.078 $mmol_{CO_2}/mmol_{photon}$, which is within the range reported for *Q. suber* (0.051 – 0.089
239 $mmol_{CO_2}/mmol_{photon}$) at different light conditions and times of the year (Vaz *et al.*, 2010, 2011).
240 Photon yield was also assessed in photorespiration by varying the ratio of carboxylation/oxygenation
241 activity of Rubisco using nitrate and ammonia as nitrogen source (Supplementary File 5 Fig. S2). The
242 results show that the model is able to predict a lower photosynthetic efficiency in drought conditions
243 (carboxylation/oxygenation < 3.0), and when using nitrate as nitrogen source. *In silico* simulations
244 predict an assimilation quotient of 0.76 and 0.93 $mmol_{CO_2}/mmol_{O_2}$ with nitrate and ammonia as
245 nitrogen source, respectively. These values are similar to the ones predicted in other plant leaf GSM
246 models (Poolman *et al.*, 2013, 2014).

247 The inner bark and phellogen metabolic models produce all the metabolites defined in their biomass
248 composition in heterotrophic conditions (Supplementary File 5 Tables S4-S6), using sucrose as carbon
249 source and obtaining energy through the TCA and oxidative phosphorylation. As expected, these
250 models are not able to grow in photoautotrophic conditions as several genes associated with
251 photosynthesis were considered as not expressed by *troppo*, thus, the respective reactions were
252 removed.

253 ## 2.5. Diel Multi-tissue model

254 A diel multi-tissue metabolic model was generated to analyze the metabolic interactions between leaf,
255 inner bark, and phellogen at the two phases of the diel cycle: light (day) and dark (night). Common
256 pools were created to connect the different tissues for the light and dark phases. The diel multi-tissue
257 GSM model comprises 28,426 reactions and 29,489 metabolites. This model allows the uptake of
258 minerals, such as HNO_3 and H_3PO_4 through the inner bark since the root was not considered. The
259 light/dark uptake ratio of nitrate was constrained to 3:2, as suggested in other diel GSM models
260 (Cheung *et al.*, 2014; Shaw & Cheung, 2018). The exchange of oxygen and carbon dioxide is only
261 allowed in the leaves, where photons can be uptake in the light phase. FVA and pFBA were computed,
262 setting as objective function the minimization of photons uptake for photorespiratory conditions.
263 Figure 3 depicts both the relevant transport reactions between tissues, and the metabolites stored
264 between the light and dark phases. The suberin, lignin, and cell wall sugars synthesis pathways are
265 also portrayed. The detailed simulation results are available in Supplementary File 4 – Tables S2-S3.

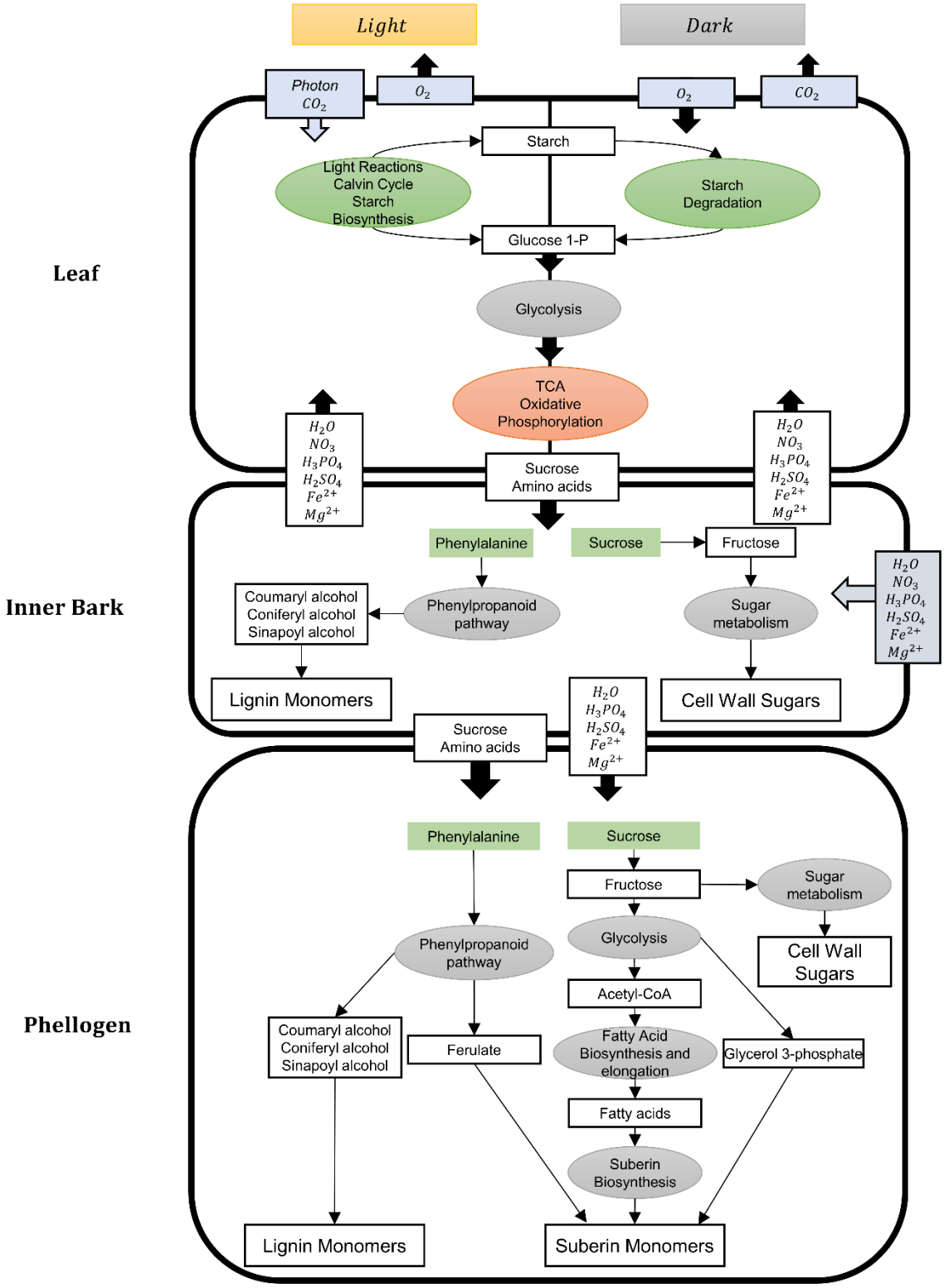


Figure 3. Schematic representation of the metabolic routes towards cork formation. Photons and gases exchanges take place in the leaf, while the uptake of inorganic ions was assumed to happen in the inner bark. Sucrose and amino acids produced in the leaf are transported into the inner bark, and then to the phellogen, where they are used for the suberin and lignin biosynthesis (besides the other biomass components). The differences in the metabolism in the day and night phases are

271 only represented for the leaf since, according with *in silico* simulations, the pathways used in the inner bark and phellogen
272 was similar in the two phases of the diel cycle. The representation is based on pFBA and FVA predictions.

273 3. Discussion

274 3.1. Comparison with other models

275 *A. thaliana* and *Z. mays* are reference organisms for C3 and C4 plants; thus, several models for these
276 organisms are available. Nevertheless, such models for less studied plants are also becoming available
277 (Grafahrend-Belau *et al.*, 2013; Yuan *et al.*, 2016b; Pfau *et al.*, 2018; Moreira *et al.*, 2019; Shaw &
278 Maurice Cheung, 2019). The iEC7871 was compared with *A. thaliana* and *Z. mays* models that provide
279 KEGG identifiers, which facilitates the comparison.
280 The cork oak model presents 1117 reactions and 945 metabolites not available in the other two
281 models. Interestingly, the annotation of the genes associated with most of these reactions was based
282 on *A. thaliana*'s homologs. The information available in Swiss-Prot for such species is exceptionally
283 high, while this information is scarce for woody plants, which may lead to wrong and missing gene
284 annotations as *Q. suber* is significantly different from *A. thaliana.* Another obstacle is the presence of
285 incomplete EC numbers. For instance, over 300 genes were initially annotated with incomplete EC
286 numbers associated with Cytochromes P450.
287 The KEGG Pathways associated with these reactions were also analyzed. The pathways with the higher
288 number of unique reactions are associated with the metabolism of xenobiotics, terpenoids, and other
289 secondary metabolites. Also, the "Tryptophan metabolism" and "Tyrosine metabolism" pathways
290 include the secondary metabolism of these amino acids, such as the quinolinic acid production,
291 dopamine, and its derivates.
292 When classified according to the main areas of the metabolism in KEGG, the pathways representing
293 'Lipid metabolism', 'aminoacid metabolism and carbohydrate metabolism', and 'Xenobiotics
294 biodegradation metabolism' were the most represented regarding reactions that are only present in
295 iEC7871.
296 The high number of reactions and metabolites only available in the cork oak model is most likely
297 associated with different reasons: the approach followed to develop the GSM model; the information
298 available at each reconstruction time; the metabolic differences between the three species; the
299 inclusion of pathways associated with secondary plant metabolism.

300 3.2. Secondary Metabolism Pathways

301 Plants produce a wide range of compounds through their secondary metabolism, whose function
302 includes defense against abiotic and biotic stress or beneficial interactions with other organisms (Isah,
303 2019). Many secondary metabolites have central roles in the pharmaceutical, cosmetics, perfume,
304 dye, and flavor industries. Despite the recent advances in the investigation of plant secondary

305 metabolism, detailed knowledge of these pathways is restricted to a few species (Isah, 2019), such as
306 *A. thaliana*, *O. sativa*, and *Z. mays*. These pathways are not always complete or available in biological
307 databases, implying an additional effort to find information regarding genes, reactions, and
308 metabolites associated with secondary plant metabolism. Although there is still much to learn about
309 these compounds and respective biosynthetic pathways, genome-scale modelling can provide insights
310 into a given organism's potential to produce secondary metabolites. The cork oak model
311 reconstructed in this work includes reactions associated with the biosynthesis and metabolism of
312 several secondary metabolites, steroids, and drugs.

313 The "Metabolism of xenobiotics by cytochrome P450" includes 38 reactions not found in the *A.*
314 *thaliana* and *Z. mays* models considered in the Results section. This pathway represents a set of
315 reactions, mostly associated with the cytochrome P450, responsible for the response to the presence
316 of toxic xenobiotics. The model includes the complete routes for the degradation of benzopyrene,
317 aflatoxin B1, and nicotine-derived xenobiotics. The cytochrome P450 family has been described being
318 highly upregulated in developing phellem tissues (Lopes *et al.*, 2020).

319 The model presents all the necessary reactions to produce jasmonic acid and its derivates, usually
320 called jasmonates, through the "alpha-Linoleic acid metabolism" pathway. These plant hormones are
321 associated with the regulation of growth and developmental processes, stomatal opening, inhibition
322 of Rubisco biosynthesis, and nitrogen and phosphorus uptake (Wasternack & Hause, 2013; Ruan *et*
323 *al.*, 2019). The "Steroid biosynthesis" pathway is essentially complete, enabling the production of
324 steroids like ergosterol, cholesterol, and stigmasterol. The enzymes associated with the synthesis of
325 calcitetrol and secalciferol (animal hormones) are the only ones whose encoding genes were not found
326 in the genome of *Q. suber*. Although these hormones are usually synthesized by animals, they were
327 already identified in a few plants (Aburjai *et al.*, 1998; Boland *et al.*, 2003; Jäpelt & Jakobsen, 2013).
328 The biosynthetic pathway for the synthesis of gibberellins is also available in the model. Although the
329 regulatory effect of hormones and steroids cannot be quantitively evaluated in stoichiometric models,
330 the presence of their biosynthetic pathways, connected to the core network, can be used as a source
331 of information regarding the potential to produce a certain hormone or steroid.

332 Although this model contains a considerable amount of information regarding secondary metabolism,
333 further curation of these pathways would improve the connectivity of the model and reduce the
334 existing gaps.

335 ### 3.3. Suberin, Lignin, and Waxes Biosynthesis

336 As mentioned before, suberin is the major component of phellogen, while its abundance is residual in
337 the leaf and inner bark (Figure 2). The phellogen also contains significant amounts of lignin and waxes.

338 The biosynthesis pathway of the monomers of these cork components in the model is represented in
339 Figure 4.

340 The synthesis of the aliphatic suberin monomers is associated with the KEGG's "Cutin, suberin, and
341 wax biosynthesis" pathway. Fatty acids are exported from the plastid and imported into the
342 endoplasmic reticulum, where these can be elongated to originate long-chain and very long-chain
343 fatty acids. Cytochromes P450, especially of the CYP86A family, are responsible for catalyzing several
344 reactions that synthesize suberin monomers. CYP86A1 and CYP704B1 are responsible for the ω-
345 hydroxylation of long-chain fatty acids, while CYP86B1 acts on very-long chain fatty acids.

346 Epoxide hydrolases (EC: 3.3.2.10), ω-oxo-fatty acid dehydrogenases (EC: 1.2.1.-), ω-hydroxy-fatty acid
347 dehydrogenases (EC: 1.1.1.-), and peroxygenases (EC: 1.11.2.3) can act in the ω-hydroxy acid in
348 different combinations, to originate the diverse α,ω-dicarboxylic, poly α,ω-dicarboxylic, and
349 polyhydroxy-fatty acids. The genome annotation and the model manual curation allowed identifying
350 38 genes encoding epoxide hydrolase, five encoding ω-hydroxy-fatty acid dehydrogenase, and six
351 encoding peroxygenases. Nevertheless, no genes encoding ω-oxo-fatty acid dehydrogenase were
352 available in any database; thus, the reactions catalyzed by this enzyme were added without gene
353 association.

354 The polymerization process is still not completely understood but it likely involves esterification
355 reactions through glycerol-3-phosphate acyltransferases (GPAT), producing acylglycerol esters. This
356 family of enzymes was identified as a key step in suberin/cork polymerization and the expression level
357 of GPAT5 gene is higher in June, which corresponds to a period of higher phellogen activity (Marum
358 *et al.*, 2011). Acylglycerol esters are then secreted through the Golgi secretory pathway and ABC
359 transporters (which are overexpressed in phellem (Lopes *et al.*, 2020) and incorporated in the suberin
360 glycerol polyester (Vishwanath *et al.*, 2015). The aliphatic suberin polyester transport and assembly
361 were simplified in the model and represented through a reaction converting the determined
362 precursors (Supplementary File 3 – Table S6) into the macromolecular representation of suberin: "e-
363 Suberin".

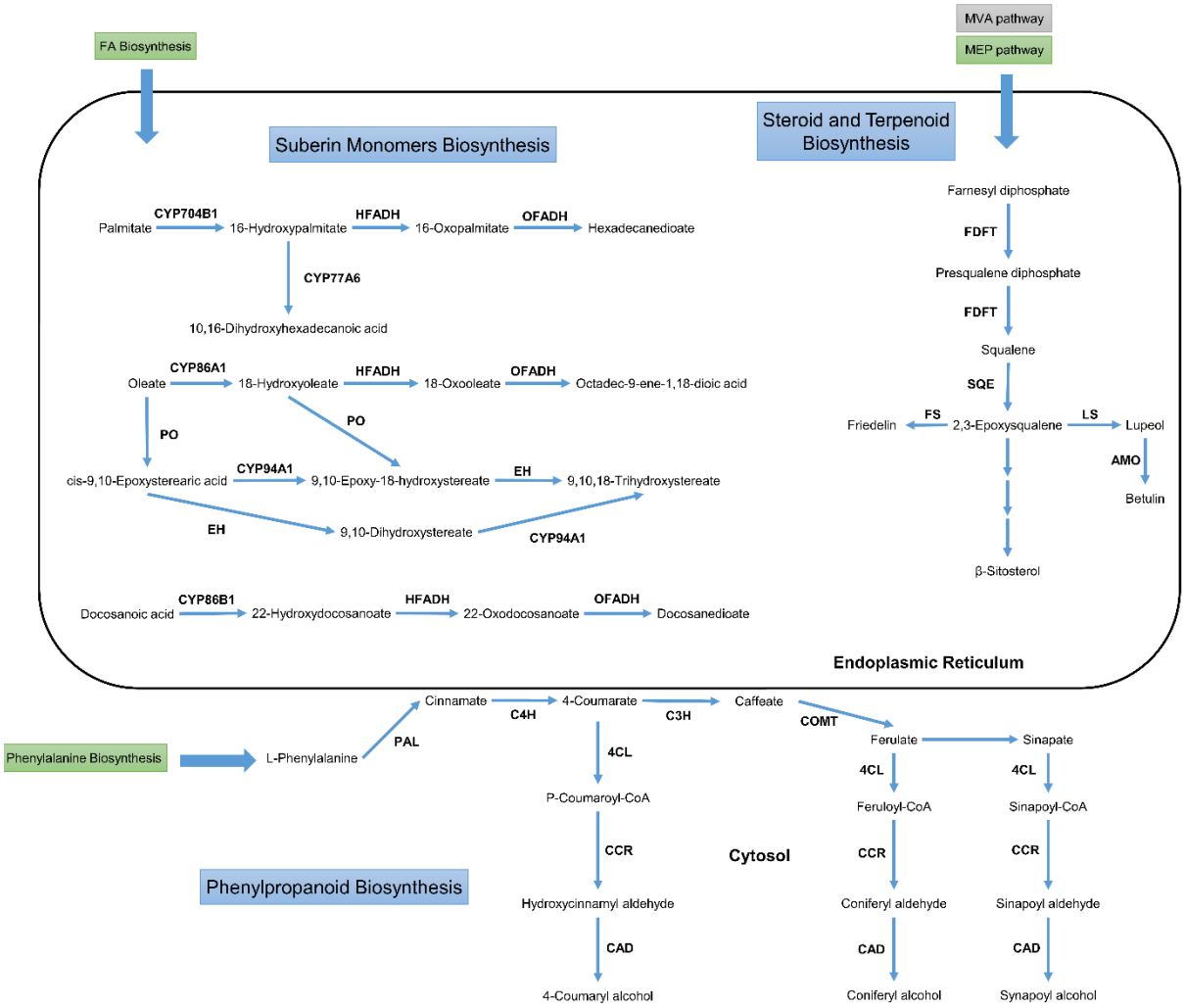


Figure 4. Biosynthetic pathway of suberin, waxes, and lignin monomers. The C:16 and C:18 fatty acids produced in the plastid are transported to the endoplasmic reticulum, where they can be elongated and unsaturated, and follow the suberin monomer biosynthesis pathway. Cytochromes P450 (CYPs), peroxygenases, epoxy hydroxylases, ω-oxo-fatty acid dehydrogenases, ω-hydroxy-fatty acid dehydrogenases catalyze successive reactions to produce the aliphatic monomers. Farnesyl diphosphate, produced from isopentenyl diphosphate provided by the cytosolic mevalonate (MVA) pathway or by the plastidic methylerythritol phosphate (EMP) pathway is used as the initial precursor of steroids in the ER. 2,3-Epoxysqualene is converted into sterols and terpenoids, monomers of the wax component of the phellogen. The phenylpropanoid pathway uses phenylalanine produced in the leaf's chloroplasts, to produce cinnamate. The pathway follows in the cytosolic surface of the endoplasmic reticulum, and then in the cytosol, producing the lignin monomers. 4CL – 4-Coumarate CoA ligase, AMO: β-amyrin monooxygenase, C3H: p-coumarate 3-hydroxylase, C4H: cinnamate 4-hydroxylase, CAD: cinnamyl alcohol dehydrogenase, CCR: cinnamoyl-CoA reductase, COMT: Caffeoyl-CoA O-methyltransferase, EH: epoxide hydroxylase, FDFT: farnesyl diphosphate farnesyltransferase, FS: friedelin synthase, HFADH: ω-hydroxy-fatty acid dehydrogenase, LS - lupeol synthase, OFADH: ω-oxo-fatty acid dehydrogenase, PAL: phenylalanine ammonia lyase, PO: peroxygenase, SQE: squalene monooxygenase. The IDs of the reactions catalyzed by these enzymes can be found in Table S10 of Supplementary File 2.

Waxes are also an important cork component and are composed essentially of sterols and terpenoids (Castola *et al.*, 2005). These are produced through the "Steroid biosynthesis" pathway in the endoplasmic reticulum. The initial precursor, farnesyl diphosphate, is synthesized in the cytosolic mevalonate pathway or by the plastidic methylerythritol phosphate pathway. Farnesyl diphosphate farnesyltransferase and squalene monooxygenase convert farnesyl diphosphate into 2,3 –

385 epoxysqualene. This triterpenoid can be converted into cyclic tripterpenoids, such as friedelin and
386 lupeol, or can be used to produce phytosterols like β-sitosterol, through a longer pathway.

387 The cork contains a considerable amount of lignin, although this polymer is present in other tissues.
388 The synthesis of the respective monomers (guaiacyl lignin, hydroxyphenyl lignin, and syringyl lignin) is
389 described in the KEGG's "Phenylpropanoid biosynthesis" pathway. Phenylalanine is the main
390 precursor of this pathway that occurs mostly in the cytoplasm. It is converted to cinnamic acid by
391 phenylalanine ammonia-ligase (EC: 4.3.1.24). Cinnamate monooxygenase, caffeate
392 methyltransferase, and ferulate-5-hydroxylase are responsible for the successive conversion of
393 cinnamic acid to coumarate, ferulate, hydroxyferulate, and sinapic acid, in the cytosolic surface of the
394 endoplasmic reticulum. These metabolites can be converted into lignin monomers in a three-step
395 process catalyzed by coumarate-CoA ligase (EC: 6.2.1.12), cinnamoyl-CoA reductase (EC: 1.2.1.44), and
396 cinnamyl-alcohol dehydrogenase (EC: 1.1.1.195). Through the polymer assembly in the cell wall,
397 peroxidases (EC: 1.11.1.7) convert the guaiacyl, hydroxyphenyl, and syringyl alcohols into guaiacyl,
398 hydroxyphenyl, and syringyl lignins, respectively. The linkages between aliphatic and aromatic suberin
399 were not included in the model since these macromolecules are represented in the biomass
400 separately.

401 ### 3.4. Tissue-specific models

402 A generic metabolic model comprises all the reactions catalyzed by all the enzymes encoded in an
403 organism's genome. However, this approach does not account for the regulatory network present in
404 each organ, tissue, or cell in different environmental conditions. The integration of transcriptomics
405 data in a GSM model allows obtaining metabolic models closer to the *in vivo* phenotype of the
406 respective tissue or condition.
407 Transcriptomics data for the leaf, inner bark, and phellogen were integrated into the generic model,
408 originating three tissue-specific models. The biomass of each tissue was determined using data
409 available in the literature and other metabolic models. The leaf macromolecular composition was
410 mostly based on previously published GSM models. Quantitative information regarding the leaf
411 composition of *Q. suber* would be useful to increase the model's reliability and improve predictions.
412 The composition of the inner bark and the phellogen was based on experimental data available for *Q.*
413 *suber*. While the inner bark is mostly composed of lignin and carbohydrates, the phellogen also
414 comprises suberin and waxes. Suberin is composed of glycerol, ferulic acid, and diverse alkanols, fatty
415 acids, hydroxy acids, ω-hydroxy acids and α,ω-dicarboxylic acids. Waxes are composed of terpenes,
416 such as friedelin, and sterols, like β-sitosterol. Although the cork bark contains tannins (mostly

417 ellagitannins) (Cadahía *et al.*, 1998), these compounds were not included in the biomass since their
418 biosynthetic pathways are poorly understood and not available in the used biological databases.
419 The number of reactions and metabolites on each subcellular compartment was compared across the
420 three tissues. The number of transport reactions in the Reproduction Phellogen is visibly higher than
421 in the other tissues, including the Virgin Phellogen. These reactions are essentially associated with the
422 transport of steroids and hormones. Based on BLAST searches against TCDB, the genes associated with
423 these reactions are auxin (phytohormone required for cell division (Perrot-Rechenmann & Napier,
424 2005)) efflux pumps. An analysis of the pathways available in the Reproduction and Virgin Phellogen
425 models, followed by *in silico* simulations did not allow to identify significant differences between
426 them. The cork oak phellogen's transcriptional profile allows inferring that the flavonoid route is
427 favored in bad quality cork, while the lignin and suberin production pathways are preferred in good
428 quality cork production (Teixeira *et al.*, 2018). A recent study reported a higher expression of genes
429 associated with fatty acid biosynthesis and elongation in reproduction cork (Lopes *et al.*, 2020).
430 Although it was reported that the macromolecular composition of reproduction and virgin cork could
431 be similar, the precursors of suberin and extractives are significantly different (Pereira, 1988). Hence,
432 a biomass composition specific to the virgin phellogen should be defined, including the tannin content,
433 to allow the observation of metabolic differences between the two models.

434 ### 3.5. Diel multi-tissue model

435 Tissues and organs of multicellular organisms do not perform their metabolic functions individually.
436 Instead, they interact with each other by exchanging sugars, amino acids, hormones, and others. In
437 plants, the simplest example is the sucrose formation in the leaves and its transport across all other
438 organism tissues. Multi-tissue GSM models allow analyzing the dependencies between tissues in
439 terms of biomass precursors, carbon skeletons, nitrogen, sulfur and phosphorus sources, and the
440 energy required for their translocation and proton balance (Bordbar *et al.*, 2011; Zakhartsev *et al.*,
441 2016). The introduction of light and dark phases allows predicting features that are not possible in
442 continuous light models, such as the accumulation and utilization of carboxylic acids (Cheung *et al.*,
443 2014).
444 The leaf, inner bark, and phellogen models were merged into a multi-tissue GSM model. Light and
445 dark phases were considered to present the metabolic differences that occur in these two phases.
446 Other tissues or organs, such as the root and xylem, could also be considered and included in the
447 multi-tissue model. Nevertheless, this work's scope is related to the metabolism of cork precursors
448 and including more tissues would increase the complexity of the model unnecessarily.

449 The most significant differences between the light and dark phases of metabolism were identified in
450 the leaf (Supplementary File 4 – Tables S8-S9), due to the photosynthesis and photorespiration
451 pathways.
452 In the light dependent reactions, carbon dioxide is fixed through the Calvin Cycle, forming
453 carbohydrates, and consequently, the carbon skeletons for the synthesis of the remaining biomass
454 components. A fraction of the carbohydrates produced here (mainly starch) is stored and used as
455 energy source when needed. The storage of starch is more efficient in the model than sucrose,
456 glucose, or fructose. Whereas starch is mobilized in plastids, soluble sugars are accumulated in the
457 vacuole, implying an energetic cost to its transport. The model produced malate in the dark (light
458 independent reactions), which is used at day (light dependent) by the malic enzyme to produce
459 pyruvate and obtain reducing power (NADPH). Citrate was also produced at night (light independent
460 reaction) through the tricarboxylic cycle (TCA) and used to feed the diurnal TCA and provide carbon
461 skeletons for the amino acid metabolism. This behavior is in agreement with the previously published
462 diel GSM model for *A. thaliana* (Cheung *et al.*, 2014).
463 Nitrate provided by the inner bark is transported into the leaf and is then converted into ammonia. In
464 the plastids, ammonia is used to incorporate nitrogen through the glutamine synthetase-glutamate
465 synthase pathway.
466 Sucrose and amino acids are exported from the leaf and then imported by proton symport by the inner
467 bark. The proton balance of the common pool is maintained by plasma membrane ATPases. In the
468 inner bark, sucrose is used as an energy and carbon source to produce the biomass components. As
469 mentioned before, the inner bark biomass is mainly composed of carbohydrates, lignin, and protein.
470 Hence, pathways associated with the synthesis of cell wall precursors ("Amino sugar and amino
471 nucleotide metabolism", "Starch and sucrose metabolism", and "Phenylpropanoid biosynthesis") are
472 the most relevant in this tissue.
473 The remaining sucrose and amino acids produced in the leaves are transported to the phellogen. The
474 reversible sucrose synthase activity allows sucrose conversion into UDP-glucose, which is used to
475 produce other cell wall components, and fructose that follows the glycolytic pathway to produce
476 pyruvate. A fraction of it is transported into the plastid and converted by pyruvate dehydrogenase
477 into acetyl-CoA, the fatty acids precursor. As mentioned above, fatty acids are elongated and suffer a
478 series of hydroxylation, epoxidation, and peroxidation reactions in the endoplasmic reticulum.
479 Ferulate, another precursor of suberin, is produced through the "Phenylpropanoid biosynthesis"
480 pathway, with phenylalanine serving as a precursor. The glycerol 3-phosphate used for the
481 esterification reactions is produced from glycerone phosphate produced through glycolysis.

482 The diel multi-tissue GSM model developed in this work is a useful framework capable of providing
483 insights into the metabolism of *Q. suber* at a global level. It can be used to study the biosynthetic
484 pathways of suberin, lignin, waxes, and many other compounds.

485 Conclusions

486 In this work, we present the first genome-scale metabolic model of a woody plant. The iEC7871 was
487 based on knowledge retrieved from genomic information, biological databases, and literature. It can
488 simulate the Cork Oak tree's metabolic behavior in phototrophic and heterotrophic conditions, as well
489 as photorespiration. This model comprises the pathways of the central metabolism and several
490 pathways associated with the secondary metabolism reproducing the formation of the major
491 components of cork.

492 The integration of transcriptomics data allowed obtaining tissue-specific models for the leaf, inner
493 bark, and phellogen. These models were merged to obtain a multi-tissue GSM model that comprises
494 the diel cycle's dark and light phases.

495 This GSM model comprehends the four main secondary metabolic pathways participating in cork
496 production: acyl-lipids, phenylpropanoids, isoprenoids, and flavonoids. The lipid biosynthesis pathway
497 is required for the biosynthesis of the linear long-chain compounds forming the aliphatic suberin
498 domain, which share upstream reactions with waxes biosynthesis. The phenylpropanoid metabolism
499 is needed for the biosynthesis of the cork aromatic components, which share reactions with wood
500 lignin.

501 The metabolic models developed in this work can be used as a tool to analyze and predict the
502 metabolic behavior of the tree and evaluate its metabolic potential. Metabolic modelling methods can
503 be applied, including dynamic approaches, to study the changes in this tree's metabolism over time
504 and environment.

505 4. Materials and Methods

506 4.1. Software

507 *merlin* v4 was used to support the reconstruction process, while COBRApy v0.20.0 (Ebrahim *et al.*,
508 2013) was used to perform all simulations and analyses of the GSM model, as well as generate the diel
509 multi-tissue model. The simulations were performed using the CPLEX v128.0.0 solver.
510 The Troppo (Ferreira *et al.*, 2020) python package was used to integrate the transcriptomics data in
511 the metabolic model, originating tissue-specific models.

512 FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), Sickle (Joshi & Fass, 2011),
513 Bowtie2 (Langmead & Salzberg, 2012), FeatureCounts (Liao *et al.*, 2014), and edgeR (Robinson *et al.*,
514 2009) were used to process the transcriptomics data.

515 ## 4.2. Metabolic Reconstruction

516 The "Automatic workflow" tool, available in *merlin*, was used to perform the genome annotation by
517 assigning EC numbers to enzyme encoding genes based on BLAST searches against Swiss-Prot and
518 TrEMBL.
519 A draft metabolic network was assembled by loading KEGG's metabolic information and integrating
520 the genome annotation. KEGG reactions associated with enzymes identified in the genome annotation
521 stage were included in the model, as well as spontaneous reactions.
522 Transport reactions were automatically generated using the TranSyT (Lagoa *et al.*, 2021) tool.
523 Nevertheless, additional transport reactions were added if reported in the literature, or if necessary
524 for the model functionality. The subcellular location of proteins was predicted using the WolfPsort
525 tool. Additionally, LocTree3 (Goldberg *et al.*, 2014) and ChloroP 1.1 (Emanuelsson *et al.*, 1999) were
526 used to verify protein locations during the manual curation stage.
527 The leaf, inner bark, and phellogen biomass compositions were based on previously published plant
528 GSMMs and available literature. The biomass precursors were organized in macromolecules or cell
529 structures, labelled as "e-Metabolites". The leaf macromolecular contents were determined using *A.*
530 *thaliana* models (Dal'Molin *et al.*, 2010; Arnold & Nikoloski, 2014). The cell wall sugar content was
531 included in the e-Carbohydrate composition, while lignin was included in the e-Lignin composition.
532 The monomer contents of DNA, RNA, and protein were determined using the biomass tool, available
533 in *merlin*. The fatty acid, lipid, and carbohydrate compositions were determined using experimental
534 data for *Q. suber* or closely related organisms when species-specific data were not available (Koiwai
535 *et al.*, 1983; Nouairi *et al.*, 2006; Passarinho *et al.*, 2006). The lignin, carbohydrate, suberin, and wax
536 contents and composition in the inner bark and phellogen were determined using available
537 experimental data (Pereira, 1988; Lourenço *et al.*, 2016). Suberin was only accounted for in the inner
538 bark and phellogen. The e-Cofactor component includes a set of universal cofactors and vitamins
539 (Xavier *et al.*, 2017), which were included in the biomass of each tissue. Nevertheless, the leaf "e-
540 Cofactor" reaction also comprises a set of pigments, such as chlorophylls and carotenoids, determined
541 according to experimental data (Garcia-Plazaola, 1997). The energy requirements were inferred as
542 reported by (Yuan *et al.*, 2016a).
543 Through the whole process, literature, and biological databases, namely KEGG, MetaCyc (Caspi *et al.*,
544 2020), and BRENDA (Chang *et al.*, 2021), were consulted to retrieve information regarding metabolic,
545 genomic, and enzymatic information, as described previously (Thiele & Palsson, 2010; Dias *et al.*,

546 2019). Briefly, KEGG Pathways allowed identifying reactions that were not included through the
547 automatic network assembly; MetaCyc was used to identify reactions not available in KEGG; MetaCyc
548 and BRENDA provided information regarding the reaction's reversibility and mass-balance.
549 The validation of the GSM model started by assuring the biomass formation, which was performed
550 using the BioISO plug-in available in *merlin*.
551 The following validation approaches were applied to the model:
552 1. Growth without photons, carbon, nitrogen, phosphorus, and sulfur sources.
553 2. Futile cycles and stoichiometrically balanced cycles.
554 3. Growth rate assessment.
555 4. Growth with different elemental sources.
556 5. Capacity to present flux through photosynthesis, respiration, and photorespiration.

557 The model validation was performed using FBA (Orth *et al.*, 2010), pFBA, and FVA, available in
558 COBRApy. The general GSM model of *Q. suber* is available in SBML format in Supplementary File 6. All
559 the metabolic models developed in this work are available at https://cutt.ly/quercussubermodels.

560 4.3. Tissue-Specific and multi-tissue Models

561 Transcriptomics data from different tissues were used to obtain tissue-specific models. The quality of
562 the transcriptomics data was analyzed using the FastQC software. After trimming the data with Sickle,
563 Bowtie2 was used to align the trimmed FASTQ files against the reference genome. FeatureCounts was
564 used to count the mapped reads to each gene. After filtering and normalizing the data with the edgeR
565 library, datasets with the normalized counts of each gene were obtained. These files were used as
566 input for *troppo*, together with the generic GSMM, to generate tissue-specific models. *Troppo* was run
567 using the CPLEX solver and the fastcore reconstruction algorithm. A *troppo*'s integration strategy was
568 developed to guarantee that the tissue-specific models are capable of producing biomass. The median
569 of each dataset was used as threshold, and the remaining parameters were kept as default. *Troppo*
570 identifies the reactions that should be removed from the generic model, based on the expression of
571 the genes included in the gene-protein-reaction association of each reaction. Nevertheless, the
572 reactions that were maintained in the model within this approach kept all of their respective genes.
573 The carotenoid biosynthesis has two branches, one associated with trans-Phytofluene and the other
574 with 15,9-dicis-Phytofluene. The enzymes present in the two branches are similar (phytoene
575 desaturase and zeta-carotene desaturase). However, one of them is dependent on photosynthesis
576 since it unbalances the plastoquinone/plastoquinol ratio. The other one is stoichiometrically balanced
577 since it involves successive oxidation and reduction of these quinones. Hence, the presence of the

578 second branch was guaranteed in the leaf model since it is essential for the carotenoid production
579 through the dark phase.
580 The tissue-specific models obtained using *troppo* were adapted to account for the light and dark
581 conditions, by duplicating all reactions and metabolites. Reactions converting sugars (starch, sucrose,
582 glucose, fructose, malate, citrate, quercitol, and quinate), 18 amino acids, and nitrate between light
583 and dark were added to the model, based on a previously published approach (Cheung *et al.*, 2014).
584 To build the multi-tissue model, the diel tissue-specific models were merged, and common pools
585 between leaf and inner bark (common pool 1), and inner bark and phellogen (common pool 2) were
586 created. Since the root was not accounted for, the uptake of water, nitrogen, sulfur, phosphorus, and
587 magnesium sources was included in the inner bark.

588 ## 4.4. Simulations

589 Two different approaches were used to perform simulations with pFBA. The first was to fix the biomass
590 formation to 0.11 h^{-1}, and set the minimization of the photon uptake as the objective function. In the
591 second approach, the photon uptake was fixed to 100 $mmol_{photon}.gDW^{-1}.h^{-1}$, maximizing the
592 biomass production. Photorespiration was simulated by constraining the carboxylation/oxygenation
593 flux ratio (Vc/Vo) of Rubisco (reactions R00024 and R03140) to 3:1 (Weber, 2007). Simulations with
594 the inner bark, and virgin and reproduction phellogen were performed by maximizing the biomass
595 production while limiting the sucrose and amino acid uptake to 1 $mmol_{photon}.gDW^{-1}.h^{-1}$.

596 The simulations with the diel multi-tissue GSM model were performed using the restrictions applied
597 in photorespiration, with the additional constraint of the nitrate light/dark uptake ratio, which was
598 settled to 3:2 (Cheung *et al.*, 2014; Shaw & Cheung, 2018).

599 ## 4.5. Accession Numbers

600 The genome sequence of *Quercus suber* was retrieved from the NCBI database with the assembly
601 accession number GCF_002906115.1. The transcriptomics data was retrieved from the EBI database(Li
602 *et al.*, 2015) (Accession PRJNA392919 for Leaf and Inner Bark; Accession PRJEB33874 for Phellogen).

5. Supporting Information

5.1. Supplementary file 1: Results of the genome annotation, and respective analysis.

5.2. Supplementary file 2: Reactions and metabolites with KEGG IDs, included in the cork oak, A. thaliana, and Z. mays models (S1-S5). Analysis of the pathways whose reactions were only identified in the cork oak model according with the genome annotation (S6), and by tissue (S7). Distribution of reactions by pathway in each tissue (S8).

5.3. Supplementary file 3: Biomass Composition. Data collected from publications, and respective treatment to include it in the GSM models.

5.4. Supplementary file 4: Flux values, maximum, and minimum according to with pFBA and FVA simulations for the reproduction and virgin phellogen models, and the transport reactions across tissues and storage reactions in the diel multi-tissue model.

5.5. Supplementary file 5: Validation of the metabolic models.

5.6. Supplementary file 6: iEC7871 model in SBML format.

6. Acknowledgments and Funding

The authors would like to acknowledge project 22231/01/SAICT/2016: “Biodata.pt – Infraestrutura Portuguesa de Dados Biológicos”, supported by Lisboa Portugal Regional Operational Programme (Lisboa2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). The authors would also like to acknowledge the Portuguese Foundation for Science and Technology (FCT) for the Strategic Funding FCT 2020-2023 (PEst UIDB/04469/2020). E. Cunha acknowledges FCT for the scholarship (DFA/BD/8076/2020). Oscar Dias acknowledges FCT for the Assistant Research contract obtained under CEEC Individual 2018. Inês Chaves was funded by DL 57/2016/CP1351/CT0003.

7. Author Contributions

E.C. and M.S. developed and tested the metabolic models; I.C. analyzed the data regarding biochemical and molecular biology aspects of the cork oak; H.D. and D.L. processed the transcriptomics data; D.R.L. developed methods for data processing and analysis, and model reconstruction; E.C. wrote the original draft; I.C., M.R., I.R., and O.D. reviewed and edited the article; M.R., I.R., and O.D. conceived and supervised the project.

8. Data Availability

All metabolic models developed in this work are available at https://bit.ly/38Iez8h.

634 **References**

635 **Aburjai T, Al-Khalil S, Abuirjeie M**. **1998**. Vitamin D3 and its metabolites in tomato, potato, egg plant
636 and zucchini leaves. *Phytochemistry*.

637 **Acácio V, Holmgren M, Jansen PA, Schrotter O**. **2007**. Multiple recruitment limitation causes arrested
638 succession in Mediterranean cork oak systems. *Ecosystems*.

639 **Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J**. **2014**. Identification of anticancer
640 drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling.
641 *Molecular Systems Biology*.

642 **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. **1990**. Basic local alignment search tool. *Journal*
643 *of Molecular Biology* **215**: 403–410.

644 **Apweiler R, Martin MJ, O'Donovan C, Magrane M, Alam-Faruque Y, Antunes R, Barrell D, Bely B,**
645 **Bingley M, Binns D, *et al.* 2011**. Ongoing and future developments at the Universal Protein Resource.
646 *Nucleic Acids Research* **39**: D214–D219.

647 **Arnold A, Nikoloski Z**. **2014**. Bottom-up Metabolic Reconstruction of Arabidopsis and Its Application
648 to Determining the Metabolic Costs of Enzyme Production. *Plant Physiology* **165**: 1380–1391.

649 **Boland R, Skliar M, Curino A, Milanesi L**. **2003**. Vitamin D compounds in plants. *Plant Science*.

650 **Bordbar A, Feist AM, Usaite-Black R, Woodcock J, Palsson BO, Famili I**. **2011**. A multi-tissue type
651 genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Systems*
652 *Biology*.

653 **Cadahía E, Conde E, Fernández De Simón B, García-Vallejo MC**. **1998**. Changes in Tannic Composition
654 of Reproduction Cork Quercus suber throughout Industrial Processing. *Journal of Agricultural and*
655 *Food Chemistry*.

656 **Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S,**
657 **Subhraveti P, Karp PD**. **2020**. The MetaCyc database of metabolic pathways and enzymes-a 2019
658 update. *Nucleic Acids Research*.

659 **Castola V, Marongiu B, Bighelli A, Floris C, Laï A, Casanova J**. **2005**. Extractives of cork (Quercus suber
660 L.): chemical composition of dichloromethane and supercritical CO2 extracts. *Industrial Crops and*
661 *Products* **21**: 65–69.

662 **Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, Neumann-Schaal M, Jahn D,**
663 **Schomburg D**. **2021**. BRENDA, the ELIXIR core data resource in 2021: New developments and updates.

Nucleic Acids Research.

Cheung CYM, Poolman MG, Fell DA, Ratcliffe RG, Sweetlove LJ. **2014**. A Diel Flux Balance Model Captures Interactions between Light and Dark Metabolism during Day-Night Cycles in C 3 and Crassulacean Acid Metabolism Leaves. *Plant Physiology* **165**: 917–929.

Cheung CYM, Williams TCR, Poolman MG, Fell DA, Ratcliffe RG, Sweetlove LJ. **2013**. A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *The Plant journal : for cell and molecular biology* **75**: 1050–61.

Chung BKS, Lakshmanan M, Klement M, Mohanty B, Lee DY. **2013**. Genome-scale in silico modeling and analysis for designing synthetic terpenoid-producing microbial cell factories. *Chemical Engineering Science*.

Dal'Molin CG de O, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. **2010**. AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiology*.

Dias O, Rocha I. **2015**. Systems Biology in Fungi. In: Paterson R, ed. Molecular Biology of Food and Water Borne Mycotoxigenic and Mycotic Fungi. Boca Raton: CRC Press, 69–92.

Dias O, Rocha M, Ferreira EC, Rocha I. **2015**. Reconstructing genome-scale metabolic models with merlin. *Nucleic Acids Research* **43**: 3899–3910.

Dias O, Saraiva J, Faria C, Ramirez M, Pinto F, Rocha I. **2019**. iDS372, a Phenotypically Reconciled Model for the Metabolism of Streptococcus pneumoniae Strain R6. *Frontiers in Microbiology* **10**.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BØ. **2007**. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 1777–82.

Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. **2013**. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology* **7**.

Emanuelsson O, Nielsen H, Heijne G Von. **1999**. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**: 978–984.

Ferreira J, Vieira V, Gomes J, Correia S, Rocha M. **2020**. *Troppo* - A Python Framework for the Reconstruction of Context-Specific Metabolic Models. In: Advances in Intelligent Systems and Computing. Springer Verlag, 146–153.

Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. **2005**. In silico

694 design and adaptive evolution of Escherichia coli for production of lactic acid. *Biotechnology and*
695 *Bioengineering*.

696 **Garcia-Plazaola J**. **1997**. Seasonal changes in xanthophyll composition and photosynthesis of cork oak
697 (Quercus suber L.) leaves under Mediterranean climate. *Journal of Experimental Botany*.

698 **Goldberg T, Hecht M, Hamp T, Karl T, Yachdav G, Ahmed N, Altermann U, Angerer P, Ansorge S,**
699 **Balasz K, *et al.* 2014**. LocTree3 prediction of localization. *Nucleic Acids Research* **42**: W350.

700 **Graça J**. **2015**. Suberin: The biopolyester at the frontier of plants. *Frontiers in Chemistry*.

701 **Graça J, Pereira H**. **2004**. The periderm development in Quercus suber. *IAWA Journal*.

702 **Grafahrend-Belau E, Junker A, Eschenröder A, Müller J, Schreiber F, Junker BH**. **2013**. Multiscale
703 metabolic modeling: Dynamic flux balance analysis on a whole-plant scale. *Plant Physiology*.

704 **Isah T**. **2019**. Stress and defense responses in plant secondary metabolites production. *Biological*
705 *Research* **52**: 39.

706 **Jäpelt RB, Jakobsen J**. **2013**. Vitamin D in plants: A review of occurrence, analysis, and biosynthesis.
707 *Frontiers in Plant Science*.

708 **Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, Seashore-Ludlow B,**
709 **Weinstock A, Geiger T, Clemons PA, *et al.* 2014**. Predicting cancer-specific vulnerability via data-
710 driven detection of synthetic lethality. *Cell*.

711 **Joshi N, Fass J**. **2011**. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. :
712 Available at https://github.com/najoshi/sickle.

713 **Kanehisa M, Goto S**. **2000**. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*.

714 **Kim HN, Jin HY, Kwak MJ, Khaine I, You HN, Lee TY, Ahn TH, Woo SY**. **2017**. Why does Quercus suber
715 species decline in Mediterranean areas? *Journal of Asia-Pacific Biodiversity*.

716 **Koiwai A, Suzuki F, Matsuzaki T, Kawashima N**. **1983**. The fatty acid composition of seeds and leaves
717 of Nicotiana species. *Phytochemistry* **22**: 1409–1412.

718 **Lagoa D, Liu F, Faria JP, Henry CS, Dias O**. **2021**. TranSyT, an innovative framework for identifying
719 transport systems. *bioRxiv doi: 10.1101/2021.04.29.441738*.

720 **Lakshmanan M, Zhang Z, Mohanty B, Kwon J-Y, Choi H-Y, Nam H-J, Kim D-I, Lee D-Y**. **2013**. Elucidating
721 rice cell metabolism under flooding and drought stresses using flux-based modeling and analysis. *Plant*
722 *physiology* **162**: 2140–50.

723 **Langmead B, Salzberg SL**. **2012**. Fast gapped-read alignment with Bowtie 2. *Nature Methods 2012 9:4*
724 **9**: 357–359.

725 **Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G,**
726 **Purvine SO, Lopez-Ferrer D, *et al.* 2010**. Omic data from evolved E. coli are consistent with computed
727 optimal growth from genome-scale models. *Molecular systems biology* **6**: 390.

728 **Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R**. **2015**. The
729 EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*.

730 **Liao Y, Smyth GK, Shi W**. **2014**. featureCounts: an efficient general purpose program for assigning
731 sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

732 **Lopes ST, Sobral D, Costa B, Perdiguero P, Chaves I, Costa A, Miguel CM**. **2020**. Phellem versus xylem:
733 Genome-wide transcriptomic analysis reveals novel regulators of cork formation in cork oak. *Tree*
734 *Physiology*.

735 **Lourenço A, Rencoret J, Chemetova C, Gominho J, Gutiérrez A, Del Río JC, Pereira H**. **2016**. Lignin
736 composition and structure differs between Xylem, Phloem and phellem in quercus suber L. *Frontiers*
737 *in Plant Science* **7**.

738 **Marum L, Miguel A, Ricardo PC, Miguel C**. **2011**. Identification of GPAT acyltransferases in cork oak.
739 *BMC Proceedings*.

740 **McCloskey D, Palsson BØ, Feist AM**. **2013**. Basic and applied uses of genome-scale metabolic network
741 reconstructions of Escherichia coli. *Molecular systems biology* **9**: 661.

742 **Moreira TB, Shaw R, Luo X, Ganguly O, Kim HS, Coelho LGF, Cheung CYM, Williams TCR**. **2019**. A
743 genome-scale metabolic model of soybean (Glycine max) highlights metabolic fluxes in seedlings.
744 *Plant Physiology*.

745 **Nouairi I, Ammar W Ben, Youssef N Ben, Daoud DBM, Ghorbal MH, Zarrouk M**. **2006**. Comparative
746 study of cadmium effects on membrane lipid composition of Brassica juncea and Brassica napus
747 leaves. *Plant Science* **170**: 511–519.

748 **Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ**. **2011**. A comprehensive genome-
749 scale reconstruction of Escherichia coli metabolism--2011. *Molecular systems biology* **7**: 535.

750 **Orth JD, Thiele I, Palsson BØ**. **2010**. What is flux balance analysis?. *Nature Biotechnology* **28**: 245–248.

751 **Passarinho JAP, Lamosa P, Baeta JP, Santos H, Ricardo CPP**. **2006**. Annual changes in the
752 concentration of minerals and organic compounds of Quercus suber leaves. *Physiologia Plantarum*

127: 100–110.

Pereira-Leal JB, Abreu IA, Alabaça CS, Almeida M, Almeida P, Almeida T, Amorim M, Araújo S, Azevedo H, Badia A, *et al.* 2014. A comprehensive assessment of the transcriptome of cork oak (Quercus suber) through EST sequencing. *BMC Genomics* **15**: 371.

Pereira H. **1988**. Chemical composition and variability of cork from Quercus suber L. *Wood Science and Technology* **22**: 211–218.

Pereira H. **2015**. The Rationale behind Cork Properties: A Review of Structure and Chemistry. *BioResources* **10**.

Perrot-Rechenmann C, Napier RM. **2005**. Auxins. In: Vitamins and Hormones. 203–233.

Pfau T, Christian N, Masakapalli SK, Sweetlove LJ, Poolman MG, Ebenhöh O. **2018**. The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Scientific Reports*.

Poolman MG, Kundu S, Shaw R, Fell DA. **2013**. Responses to light intensity in a genome-scale model of rice metabolism. *Plant Physiology*.

Poolman MG, Kundu S, Shaw R, Fell DA. **2014**. Metabolic trade-offs between biomass synthesis and photosynthate export at different light intensities in a genome-scale metabolic model of rice. *Frontiers in Plant Science* **5**: 656.

Ramos AM, Usié A, Barbosa P, Barros PM, Capote T, Chaves I, Simões F, Abreu I, Carrasquinho I, Faro C, *et al.* 2018. The draft genome sequence of cork oak. *Scientific data* **5**: 180069.

Robinson MD, McCarthy DJ, Smyth GK. **2009**. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*.

Ruan J, Zhou Y, Zhou M, Yan J, Khurshid M, Weng W, Cheng J, Zhang K. **2019**. Jasmonic acid signaling pathway in plants. *International Journal of Molecular Sciences*.

Saha R, Suthers PF, Maranas CD. **2011**. Zea mays irs1563: A comprehensive genome-scale metabolic reconstruction of maize metabolism. *PLoS ONE*.

Seaver SMD, Bradbury LMT, Frelin O, Zarecki R, Ruppin E, Hanson AD, Henry CS. **2015**. Improved evidence-based genome-scale metabolic models for maize leaf, embryo, and endosperm. *Frontiers in Plant Science* **6**.

Shaw R, Cheung CYM. **2018**. A Dynamic Multi-Tissue Flux Balance Model Captures Carbon and Nitrogen Metabolism and Optimal Resource Partitioning During Arabidopsis Growth. *Frontiers in Plant*

Science **9**.

Shaw R, Maurice Cheung CY. **2019**. A mass and charge balanced metabolic model of Setaria viridis revealed mechanisms of proton balancing in C4 plants. *BMC Bioinformatics*.

Teixeira RT, Fortes AM, Bai H, Pinheiro C, Pereira H. **2018**. Transcriptional profiling of cork oak phellogenic cells isolated by laser microdissection. *Planta*.

Thiele I, Palsson BØ. **2010**. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols* **5**: 93–121.

Vaz M, Maroco J, Ribeiro N, Gazarini LC, Pereira JS, Chaves MM. **2011**. Leaf-level responses to light in two co-occurring Quercus (Quercus ilex and Quercus suber): leaf structure, chemical composition and photosynthesis. *Agroforestry Systems* **82**: 173–181.

Vaz M, Pereira JS, Gazarini LC, David TS, David JS, Rodrigues A, Maroco J, Chaves MM. **2010**. Drought-induced photosynthetic inhibition and autumn recovery in two Mediterranean oak species (Quercus ilex and Quercus suber).

Vishwanath SJ, Delude C, Domergue F, Rowland O. **2015**. Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier. *Plant Cell Reports*.

Wasternack C, Hause B. **2013**. Jasmonates: Biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Annals of Botany*.

Weber APM. **2007**. Synthesis, Export and Partitioning of the End Products of Photosynthesis. : 273–292.

Xavier JC, Patil KR, Rocha I. **2017**. Integration of Biomass Formulations of Genome-Scale Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in Prokaryotes. *Metabolic Engineering*.

Yuan H, Cheung CYM, Hilbers PAJ, van Riel NAW. **2016a**. Flux Balance Analysis of Plant Metabolism: The Effect of Biomass Composition and Model Structure on Model Predictions. *Frontiers in Plant Science* **7**: 537.

Yuan H, Cheung CYM, Poolman MG, Hilbers PAJ, Van Riel NAW. **2016b**. A genome-scale metabolic network reconstruction of tomato (Solanum lycopersicum L.) and its application to photorespiratory metabolism. *Plant Journal*.

Zakhartsev M, Medvedeva I, Orlov Y, Akberdin I, Krebs O, Schulze WX. **2016**. Metabolic model of

812 central carbon and energy metabolisms of growing Arabidopsis thaliana in relation to sucrose
813 translocation. *BMC Plant Biology*.

814 **Zhang C, Hua Q**. **2016**. Applications of genome-scale metabolic models in biotechnology and systems
815 medicine. *Frontiers in Physiology*.

816